Natalia Norte Fernández-Pacheco[1]

The Impact of Multimodal Ensembles on Audio-Visual Comprehension: Implementing Vodcasts in EFL Contexts

[1] Universidad de Alicante, Filología inglesa, Alicante, Spain, E-mail: natalia.norte@ua.es.
http://orcid.org/0000-0002-5191-1985.

Abstract:
The ever increasing worldwide use of Information and Communication Technologies (ICT) over the last decade, has not only contributed to the way people interact with each other, but also to how languages are taught and learnt. Language digital materials such as vodcasts bring together, in an innovative, attractive and motivating manner, diverse modes of communication which promote a multimodal approach of learning. This article explores the use of multimodal ensembles in audio-visual language learning materials designed to enhance students' comprehension. This mixed methods (qualitative and quantitative) study, comprised a multimodal analysis of two language learning vodcasts from the British Council, using ELAN as the main multimodal annotation tool. The data obtained from the multimodal transcription was relevant to describe the different orchestrations of modes contained in both vodcasts. From this data, two comprehension tests, based on the ensembles found, were developed to check how each ensemble could benefit language students. The results confirmed that EFL students' audio-visual comprehension improved when there was a greater number of orchestrated modes. These findings not only emphasise the potential of multimodal materials to improve foreign language comprehension, but also encourage teachers to adapt their methods to the pervasive digital era.

Keywords: multimodal ensembles, multimodal analysis, vodcasts, audio-visual comprehension, language teaching materials, digital tools

DOI: 10.1515/mc-2018-0002

Introduction

Currently, we are living in a society surrounded by signs which may be classified as linguistic or not, but all of them represent and communicate meaning in specific contexts. Language (spoken or written) is no longer the dominant mode which represents and communicates meaning. In contrast, language is seen as another means of communication of equal importance to the other modes such as images, gestures, and music in the process of meaning-making. Due to technological innovations and the shift from page to screen, a proliferation of new forms of discourses, which foreground a multimodal approach, has recently emerged. In general terms, the multimodal approach studies how communication works and how the process of meaning-making is produced through the combination of modes, which are influenced by social and cultural aspects (Jewitt and Kress 2003; Kress and van Leeuwen 2001; Kress 2010). This social and cultural influence is determined by the context of situation and culture (Firth 1957; Halliday and Hasan 1985 and 1989; Hymes 1974; Malinowski 1923), essential elements which help in the process of comprehension. Multimodality concentrates on how each mode interacts with others and how they are orchestrated in specific contexts to produce meaning. Moreover, it pays attention to semiotic resources available in a mode, the variety of modes and their modal affordances i. e., the possibilities a mode has of expressing and representing easily, including material, cultural and historical aspects (Kress 1993). These representations of different modes working at the same time to construct meaning in particular settings are known as multimodal ensembles. When more than one mode is used intentionally by the sign maker's interests to represent or communicate, it is said they ensemble, and meaning is inferred through their interrelation. The communicator orchestrates multimodal ensembles, based on designs, where each mode has a determined function (Kress 2010). Following Jewitt (2013), in each multimodal ensemble, a range of components are influential in the process of communication and representation. Aspects such as social context, available modes, modal affordances, the agency of an individual and technologies available are necessary to obtain a material result (see Figure 1).

Natalia Norte Fernández-Pacheco is the corresponding author.

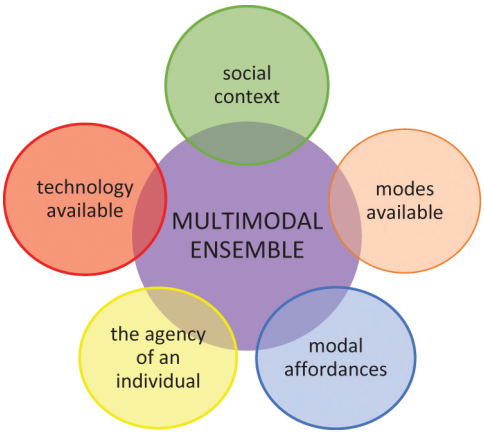

Figure 1: Elements in multimodal ensembles (based on Jewitt 2013).

In the digital age, people have become more aware of the technological affordances of multimodal resources, and consequently, their development and use has increased considerably. Discourses found in social networks (e. g., Facebook, Twitter), photo sharing (e. g., Instagram), video sharing (e. g., YouTube, UTunes) or the cross-platform mobile messaging (e. g., WhatsApp), are examples of multimodal digital tools used to create meaningful situations. The use of these new multimodal digital materials has been spread not only to everyday situations, but also to educational contexts, and especially in language learning environments. The introduction of whiteboards, computers, laptops, tablets, mp4, among other devices, in the language classroom, has promoted the use of multimodal language materials such as vodcasts. Video podcasts or vodcasts are video recordings which are uploaded to the net using Rapid Simple Syndication (RSS) feeds (Hasan and Hoon 2013). On the net we can find vodcasts related to a wide range of themes such as politics, the news, radio, television, music, fashion and beauty, literature, arts and entertainment, health, technology and education, among others, which allow language teachers to choose from a variety of topics according to their students' interests and needs. Language learning vodcasts have as their main goal to foster different aspects of language (e. g., grammar, vocabulary, pronunciation, or listening comprehension). They can be introduced as language learning activities in two different ways: creating new materials or using the available materials on the net (Rosell- Aguilar 2007). The different uses of vodcasts for language learning situations, mainly infer their use in conjunction with traditional language lessons, creating blended language learning (BLL) situations, as occurs in the study presented in this paper. The way language teachers can develop blended language learning environments concerns four different levels (Graham 2006): activity level (i. e., a learning activity encompasses face-to-face and online components), course level (i. e., an arrangement of some face-to-face and online activities within the given course), program level (i. e., the students select a fusion between face-to-face courses and online courses, or it is established in the program in advance), and institutional level (i. e., the organizational commitment to blending face-to-face instructions and online instruction). Moreover, language teachers should bear in mind some important features to determine the most suitable vodcasts to be blended in the classroom. Length, authenticity, language level, objectives within a syllabus, and engagement are some aspects to take into account in the selection (Rosell-Aguilar 2007).

The instruction and assessment of the listening skill in foreign/second language teaching has been an arduous task throughout time. In fact listening, also called the Cinderella skill (Mendelsohn 1994), was not considered to be learnt actively but as a passive process to be acquired naturally (i. e., without teaching support). In recent times, this conception has changed into a more active and strategy-based approach towards listening (Goh 2002; Goh and Taib 2006; Mendelsohn 1994; Mendelsohn and Rubin 1995; O'Malley and Chamot 1990; Oxford 1990; Rubin 1994; Vandergrift 1999, 2003, and 2008; Vandergrift and Goh 2012). The massive production and use of technological devices and multimedia language learning materials has contributed to this evolution of teaching the listening skill. Vodcasts are an example of this new multimedia language learning tools that put together different kinds of media (e. g., moving images, pictures, sounds, music, written words, and graphics) to ease comprehension. In language learning environments, multimedia learning is based on the construction of mental representations through multimodal instructional settings. For this reason, some advantages could be taken from vodcasts to improve students' foreign language audio-visual comprehension. These benefits might be founded on the cognitive theory of multimedia learning (CTML) which relates both auditory and visual inputs to facilitate comprehension and learning (Mayer 2005, 2009). CTML is grounded on the combination of three theories. Firstly, Sweller's cognitive load theory (Chandler and Sweller 1991; Sweller 1994), concerned with the use of cognitive resources while learning. Secondly, Clark and Paivio's dual coding theory (Clark and Paivio, 1991), which claims that cognition is constituted by verbal and non-verbal subsystems. And finally, Baddeley's working memory model (Baddeley 2000). According to Mayer (2005, 2009), from these theories, three

assumptions of CTML have emerged: dual-channel (i. e., people have separate channels for managing visual and auditory information), limited capacity (i. e., the amount of information we can process in the same channel at the same time is limited, and active processing (i. e., it is achieved by paying attention to important information, organizing it and creating logical mental representations to be joined to previous knowledge. Taking into account the assumptions of dual-channel, limited capacity and active processing, the CTML explains how words and pictures are perceived and understood through cognitive processes.

However, despite the technological advances and the possibility of offering students more digital materials in which visual and auditory inputs are shown together to facilitate students' comprehension, little research has been done to know their effects on language learners' audio-visual comprehension. Some studies have been carried out to demonstrate the importance of linguistic and non-linguistic knowledge to ease comprehension (e. g., Sueyoshi and Hardison 2005; Ramírez and Alonso 2007; Wagner 2010). Ramirez and Alonso (2007) demonstrated how English digital stories could be beneficial for Spanish children to better understand linguistic structures in a foreign language. Furthermore, they concluded that this kind of videos were relevant to improve not only the listening skill but others such as speaking since students were able to give some feedback, in the target language, after watching them. In the comparative study developed by Wagner (2010), it was proved how the experimental group performed a comprehension test better using an audio-visual input, in contrast to the control group which completed the same test just with audio input. In a similar study, Sueyoshi and Hardison (2005), contributed to the importance of context in terms of gestures and speakers' faces to improve comprehension. After classifying 42 low-intermediate and advanced students of English into three stimulus (audio-visual with gestures and face expression, audio-visual with no gestures or face expression, and only audio), they reported positive results, in both language levels, when audio-visual materials were shown with gestures and face expressions. However, Coniam (2001) and Suvorov (2009) have come to very dissimilar conclusions. In both, the outcomes demonstrated no significant differences between audio or video versions to perform comprehension tests. Therefore, the simultaneous use of visual and auditory material has still not been proven to be more beneficial than traditional listening comprehension activities. There is a need for more research on audio-visual comprehension, especially due to the fact that language learners are continuously being exposed to multimedia tools and multimodal texts. In this qualitative and quantitative study, I attempt to begin to fill the gap of audio-visual comprehension and multimodality, in so far as language learning is concerned.

Objectives and research questions

The main objectives of this study were to explore the role of multimodality in language learning comprehension, and more specifically, the effects on students' audio-visual comprehension when different orchestrations of modes appeared in the visualization of vodcasts. The research questions that guided this study were the following:

1. Which multimodal ensembles are found in two British Council educational vodcasts (see Figures 2 and 3) to represent and communicate meaning for language learners?
2. Does EFL students' audio-visual comprehension improve when there is a greater number of orchestrated modes?

Figure 2: Cover *English is GREAT*. https://goo.gl/C6PD1H

Figure 3: Cover *Words on the Street*. https://goo.gl/E6mCCg

Along with these research questions, I formulated the following hypothesis: EFL students' audio-visual comprehension improves when there is a greater number of orchestrated modes. As it will be revealed in the study, this hypothesis was confirmed. Taking into consideration these research questions, I designed the following study divided into two parts. In the first part, I carried out multimodal transcriptions of the two British Council vodcasts to determine which modes were used to convey meaning. These annotated transcriptions, using ELAN (see below) were analysed to establish the multimodal ensembles. In the second part, using the multimodal ensembles from the analysis, I created two tests to know the influence of the different ensembles on students' audio-visual comprehension. Thus, the study consists of first, a multimodal transcription and analysis of the vodcasts; and second, a statistical analysis of students' responses according to the number of the orchestrated multimodal ensembles.

I will begin with the first part, the multimodal transcription and analysis of the vodcasts. This is due to the fact that the findings of this analysis are needed to create the audio-visual comprehension tests of part two of the study (see Audio-visual comprehension tests)

Multimodal transcription and analysis of vodcasts

Two vodcasts from the British Council, named «English is Great» (https://goo.gl/xK4zYn) and «Camden Fashion» (https://goo.gl/4CHqrp), were chosen from the application iTunes U (free download), bearing in mind some of the characteristics established by Rosell- Aguilar (2007) such as language level, vocabulary and grammar exposure, length, entertainment and authenticity.

The preference for vodcasts from the British Council was motivated by the relevance and importance that this organization has gained throughout the years in educational environments. Moreover, these vodcasts encompassed a great variety of communicative modes which were required to fulfil the research objectives. The vodcast «English is Great» belongs to the series of vodcasts «Britain is Great», which describes some important features of the British culture such as shopping, literature, knowledge, heritage, sport, and countryside. «English is Great» (part 1), which lasts 5 minutes and 18 seconds, is set in the British library in London, and deals with the evolution of the English language. Whereas, «Camden Fashion», from the series of vodcasts «Word on the Street», shows in 4 minutes and 28 seconds, the most notable fashion styles found in this Londoner market, such as punk or Cyber-Goth.

For the purpose of this study, a multimodal transcription of the vodcasts was performed to represent the multiple communicative modes orchestrated throughout the vodcasts. The software employed for the analysis was ELAN (EUDICO Linguistic Annotator, where EUDICO stands for European Distributed Corpora Project). ELAN is a multimodal annotation tool developed by Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands, available at http://goo.gl/2CbYBU (Sloetjes and Wittenburg 2008). The diversity of modes that appeared in the vodcasts (e. g., participants' gestures, participants' speech, written language, images, and music) were classified into tiers, which made it possible to include and see together accurate annotations on the same screen.

The spoken mode was represented in the "speaker's name-transcription" tier following Norris (2004) to represent emphatic stress and Jefferson (1984) to annotate micropauses. The "speaker's name hand/arm movements" tiers depicted kinesics. Due to the limitations[1] found when trying to establish transcription conventions I developed my own list grounded on the most prominent movements. Another tier related to kinesics was "speaker's name gaze direction", annotations depicted gaze orientation (Baldry and Thibault 2006; Tan 2009)

apart from to "who" or "what" the gaze was directed towards. The "speaker's name-body position" tier represented kinesics in terms of "frontal, oblique and backside" (from the viewer's position). Following Tan's annotations (2009), a visual frame tier was included to have a general overview of the shots. The "images" tier represented the different shots appearing in the vodcasts while the speakers were not in the visual frame but they were or not heard. The music mode was shown in the "music" tier and annotations were based on three aspects. The first was embodiment, (i. e., when individuals use instruments to express themselves) or disembodiment, (i. e., when people react to the music played) (Norris 2004, p. 41). The second was loudness (Gumperz and Berenz 1993), and the third tempo (Baldry and Thibault 2006). Following the multimodal interactional analysis approach (Norris 2004, 2009), which fosters the examination of discourse in action, the analysis of both vodcasts was divided into high-level actions (i. e., "high-level action" tier), defined by Norris and Jones (2005, p.17) as «a multiplicity of connected lower-level actions». The "written" tier included the different phrases shown on the screen. Finally the "test responses" tier, (i. e., the information that the students needed to answer the test items) was included to know the specific time in which the responses to the audio-visual comprehension tests were given throughout the vodcasts, as well as the modes orchestrated. In Figures 4 and 5 we can see two screenshots from both vodcasts containing the different tiers previously depicted for the analysis.

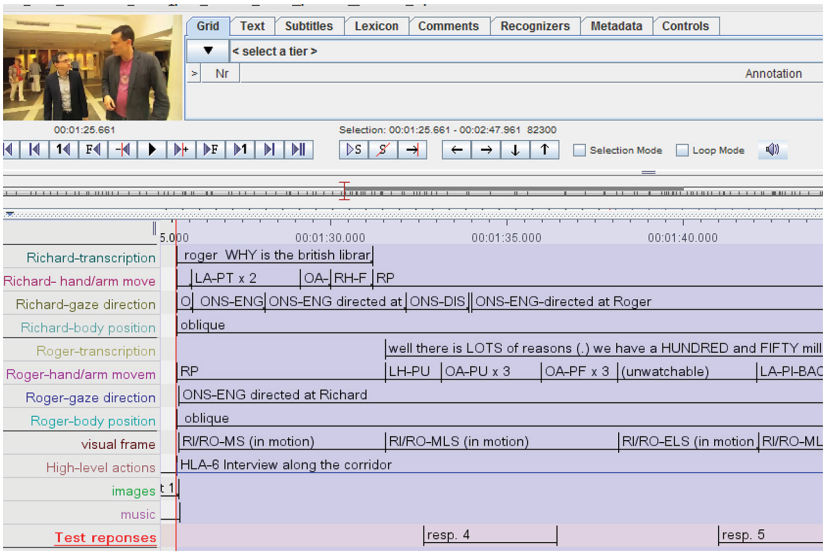

Figure 4: Tiers included in the multimodal transcription of *English is Great*.

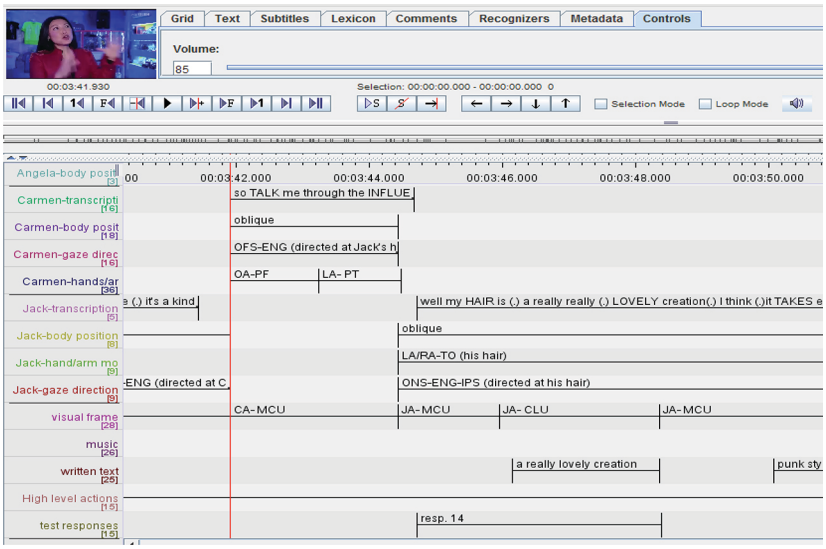

Figure 5: Tiers included in the multimodal transcription of *Camden Fashion*.

Multimodal ensembles found in the vodcasts

The previous transcription of both vodcasts facilitated the identification of different modes and the multimodal ensembles included so as to develop suitable audio-visual comprehension tests. Table 1 includes the different

multimodal ensembles found in both vodcasts. As can be observed, in the vodcast «English is Great» there was a multimodal ensemble of two modes (spoken and kinesics) and two multimodal ensembles of three modes (spoken language, kinesics, and music, and the other, spoken language, image, and music). On the other hand, in the vodcast «Camden Fashion», there were two multimodal ensembles of three modes (spoken language, kinesics and music, and the other, spoken language, kinesics and written language) and two multimodal ensembles of four modes (spoken language, written language, kinesics and music, and the other, spoken language, written language, images and music).

Table 1: Multimodal ensembles found in «English is Great» and «Camden Fashion».

Number of modes	Multimodal ensembles in «English is Great»	Multimodal ensembles in «Camden Fashion»
Two modes	spoken and kinesics	
Three modes	- Spoken language, kinesics, and music	- Spoken language, kinesics and music
	- Spoken language, image, and music	- Spoken language, kinesics and written language
Four modes		- Spoken language, written language, kinesics and music
		- Spoken language, written language, images, and music

Audio-visual comprehension tests

From the multimodal analysis two tests were designed, one for each vodcast, to check students' audio-visual comprehension (see Appendix A). These tests were given after the students had viewed each of the vodcasts (see Procedure). Two types of exercises were included in the tests, a fill in the gaps type (with 8 items), and a true/false type (with 7 items). In the case of the true/false exercise, if the item was false the students were asked to write it as a true item to avoid random responses. Although students did not know, each item from the comprehension tests was associated to a specific orchestration of modes appearing in the vodcasts (see Appendix B). On the one hand, in the vodcast «English is Great», when the responses to items 1, 2, 3, 9, and 10 were given, 2 ensembles of 3 modes (i. e., spoken language, kinesics, and music, or spoken language, image, and music) were orchestrated. However, while the responses to items 4, 5, 6, 7, 8, 11, 12, 13, 14, and 15 were given, 2 modes (i. e., spoken language and kinesics) were orchestrated. Figure 6 represents a summary of the modes orchestrated (2 modes: spoken and kinesics) when the response to item 4 was given.

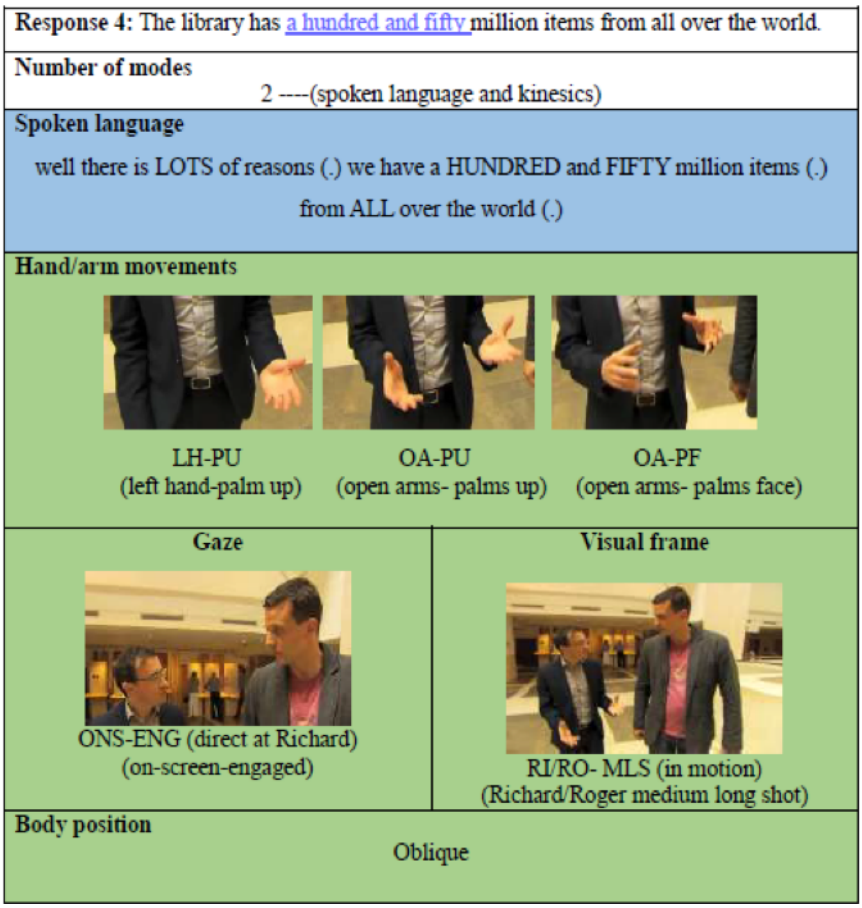

Figure 6: Two modes orchestrated in the response to item 4 in «English is Great».

On the other hand, in the vodcast «Camden Fashion», when the responses to items 1, 3, 7, 14, and 15 were given, 3 modes were orchestrated in two different ensembles (i. e., spoken language, kinesics and music, or spoken language, kinesics and written language). Whereas, four modes were orchestrated in two different multimodal ensembles (i. e., spoken language, written language, kinesics and music, or spoken language, written language, images, and music), when the responses to items 2, 4, 5, 6, 8, 9, 10, 11, 12, and 13, were given. Two examples of orchestrations of three (spoken language, kinesics and written language) and four modes (spoken language, kinesics, written language and music) are included in Figures 7 and 8 respectively.

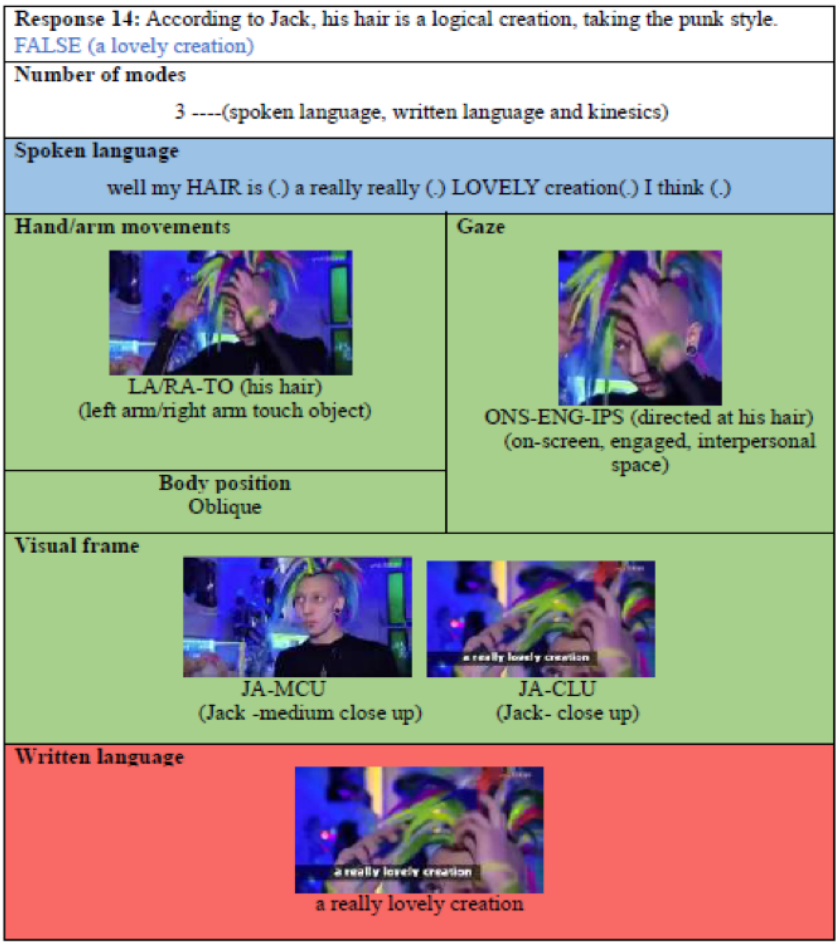

Figure 7: Three modes orchestrated in item 14 in «Camden Fashion».

Figure 8: Four modes orchestrated in item 4 in «Camden Fashion».

In total, there were 10 items associated to each orchestration of 2, 3 and 4 modes to be analysed statistically, as detailed in Table 2.

Table 2: Summary of the corresponding multimodal ensembles with the items in the tests.

Number of modes	Multimodal ensembles in«English is Great»	Items in the test «English is Great»	Multimodal ensembles in «Camden Fashion»	Items in the test «Camden Fashion»
Two modes	spoken and kinesics	4, 5, 6, 7, 8, 11, 12, 13, 14, and 15		
Three modes	- Spoken language, kinesics, and music - Spoken language, image, and music	1, 2, 3, 9, and 10	- Spoken language, kinesics and music - Spoken language, kinesics and written language	1, 3, 7, 14, and 15
Four modes			- Spoken language, written language, kinesics and music - Spoken language, written language, images, and music	2, 4, 5, 6, 8, 9, 10, 11, 12, and 13

Procedure

The 40 participants of this study were selected from a private language school in Spain. The average age among the students was between 14 and 19 years old, and the majority was female (15 males –25 female). These students were registered in the upper –intermediate or B2 level course, according to the Common European Framework of References for Languages (CEFR). They had been learning English as a foreign language for

approximately 8 to 9 years, and they had passed the Intermediate level of B1 in this language school. Despite being teenagers and «digital natives» (Prensky 2001, p.1), these students were not accustomed to using audio-visual materials to improve their listening skills. Prior to this study, students had only worked with traditional listening comprehension activities.

The study was carried out in two different days. The first day, the teacher explained to students that they were going to watch a video about the English language, and then, they had to answer some questions. Students' anxiety was reduced saying that this activity was not an exam, just a normal audio-visual exercise. After that, a comprehension test and a blank piece of paper were distributed among students. The teacher allowed students to read the test (one minute and a half), so they could have a preview of what they were going to watch. Once they had read the test, students were asked to turn it over. They were asked to turn the test over so as to avoid looking at the paper and the screen constantly. Students were allowed to take notes while visualising the vodcast, but it was not compulsory. After having given the instructions, students watched «English is Great» once. Then, they had three minutes to complete the exercises with the information they had understood. When the three minutes were over, students turned the test over and watched the vodcast again, following the same procedure previously explained. When the second visualization finished, they had three more minutes to complete the test. The second day, students watched the vodcast «Camden Fashion», following the same instructions given for «English is Great». For this vodcast, students completed a different comprehension test, but with the same characteristics as the one distributed for «English is Great».

Statistical results

The following statistical analysis was carried out to respond to the second research question, "does EFL students' audio-visual comprehension improve when there is a greater number of orchestrated modes?", and either prove or disprove my hypothesis that was the following: EFL students' audio-visual comprehension improves when there is a greater number of orchestrated modes. In order to represent the variability for each student (N = 40) on his/her audio-visual comprehension according to the modes orchestrated (2, 3 or 4 modes), I performed the following analysis. Firstly, the mean and standard deviation (SD) were calculated to provide information about students' audio-visual comprehension and the orchestration of modes as individuals and as a group. According to the means shown in Table 3, the mean comprehension score for information covered using two modes was 4.88, three modes = 5.90, and four modes 8.05. We see that there were substantial differences among them.

Table 3: Descriptive statistics table according to the students' audio-visual comprehension and the orchestrations of modes.

	Mean	Standard Deviation
two_modes	4.88	1.488
three_modes	5.90	1.646
four_modes	8.05	1.632

These differences were especially pronounced between the orchestrations of two and four modes, as can be observed in Figure 9.

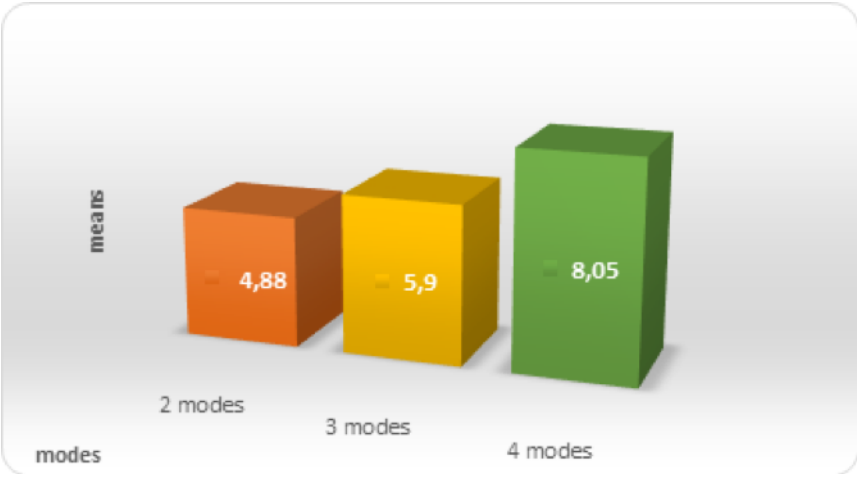

Figure 9: A comparison of means according to the orchestrations of modes presented.

Secondly, in order to examine whether these differences were statistically significant, a repeated-measures ANOVA (RM ANOVA) was conducted on the set of three scores. The partial eta squared (partial η2), was calculated to determine the amount of variance in comprehension scores that can be accounted for by mode-orchestration (i. e., condition) in Table 4.

Table 4: Test within-subjects effects (sphericity assumed).

Source Square		Type III Sum of s	df	Mean Square	F	Sig	Partial Eta Squared
factor1	Sphericity assumed	210.050	2	105.025	49.764	0.000	0.561

As we can appreciate from the table, in the analysis of tests within-subjects effects, the partial eta squared obtained was 0.56. It means that the percent of variance in the dependent variable is quite large and a positive result confirms the main hypothesis, which was that students would gain better audio-visual comprehension results when the multimodal ensembles were constituted by a major number of modes orchestrated. After carrying out the analysis to check the variability in each condition and demonstrating that it is quite high, I continued the analysis by conducting three post hoc t-tests. The purpose of this phase of the analysis was to determine whether and to what extent the three different pairs of conditions (2–3, 2–4, 3–4) differ from each other. Three paired sample t-tests were used to make post hoc comparisons between conditions (Table 5). A first paired sample indicated that there was a significant difference between two modes and three modes, t (39) = −3.537, p = 0.001. A second paired sample t-test indicated that there was a significant difference between two modes and four modes, t (39) = −8.537, p = < 0.001. The third paired sample t-test indicated that there was a significant difference between three modes and four modes, t (39) = −7.002, p = < 0.001. All pairs had a significant difference at the p < 0.05 level. Table 5 shows the paired sample tests and the results obtained.

Table 5: Paired samples test.

		Paired Differences				t	df	Sig. (2-tailed)
		Mean	Std. Deviation	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	two_modes - three_modes	−1.025	1.833	−1.611	−.439	−3.537	39	.001
Pair 2	two_modes - four_modes	−3.175	2.352	−3.927	−2.423	−8.537	39	.000
Pair 3	three_modes - four_modes	−2.150	1.942	−2.771	−1.529	−7.002	39	.000

In an attempt to know how different the conditions were, I also calculated the d effect size for each pair of scores. This effect size expresses the difference in pairs of mean scores expressed in SD units. In pair 1 (two-three modes) the d value was 0.65, in pair 2 (two-four modes) d = 1.775, and finally in pair 3 (three-four modes) d = 1.312. I followed the benchmarks of small (d = 0.40), medium (d = 0.70), and large (d = 1.00) determined by Plonsky and Osward (2014), to interpret the results. If we pay attention to these values, we can see that the differences in pair 2 and 3 are quite large. Therefore, when more modes are orchestrated, students' audio-visual comprehension improves. In other words, there is a linear relationship between more modes and more comprehension. Table 6 represents the t-test dependent results and d values.

Table 6: T-test dependent results and d values.

Pair 1 Two-three modes	t-value = −3.537 n (pairs) = 40 r for paired values = 0.319	d = 0.65
Pair 2 Two-four modes	t-value = −8.537 n (pairs) = 40 r for paired values = −0.135	d = 1.775
Pair 3 Three-four modes	t-value = −7.002	d = 1.312

n (pairs) = 40
r for paired values = 0.298

Discussion and conclusions

This study has attempted to make a small contribution to the field of audio-visual comprehension and the use of multimodal digital tools in EFL contexts. As indicated in the statistical analysis, the greater number of modes (verbal and non-verbal) correlates with the higher number of correct answers in the tests. That is to say, the greater number of modes represented in vodcasts, the better the audio-visual comprehension. It appears to be that when more modes are used, or when the information needed to respond is represented in different modes, more students with diverse learning styles are able to respond correctly. Although the two different multimodal ensembles of four modes were not statistically analysed, there was no significant difference between the numbers of correct answers when spoken language, written language, kinesics and music, or when spoken language, written language, images and music were orchestrated in the vodcasts. From these results, we can observe the influence of the three assumptions of CTML: dual-channel (i. e., people have separate channels for managing visual and auditory information), limited capacity (i. e., the amount of information we can process in the same channel at the same time is limited, and active processing (i. e., it is achieved by paying attention to important information, organizing it and creating logical mental representations to be joined to previous knowledge (Mayer 2005, 2009) and how important it is to receive auditory and visual inputs to better understand and learn information. As happened in previous research (Sueyoshi and Hardison 2005; Ramírez and Alonso 2007; Wagner 2010), this study provides positive findings towards the effects of audio-visual materials to improve second/foreign language comprehension. Consequently, it can be inferred that multimodal digital tools (e. g., vodcasts) may contribute in the process of audio-visual comprehension when learning foreign languages. Furthermore, this study provides a link between the multimodal approach, in which different communicative modes (linguistic and non-linguistic) are orchestrated to create meaningful situations, and audio-visual comprehension. Since the application of vodcasts to improve audio-visual comprehension has been developed in an educational context some pedagogical implications could also be inferred. Firstly, the use of audio-visual materials, such as vodcasts, to improve students' listening skills in foreign language contexts, might not only be more appealing for students than traditional audio tracks, but also more beneficial for students with different learning styles. Thus, the design of foreign language materials with a range of communicative modes for listening comprehension purposes might help students with specific intelligences, and facilitate students' understanding and learning. Secondly, present day representation and communication of meaning is highly multimodal; therefore, teachers should not only be concerned with written and spoken language in the materials they provide to students and/or the activities they require students to produce, but also with body language and non-verbal materials (Morell, 2018). Finally, the use of vodcasts in the classroom may help students to develop other skills such as speaking or writing. After the visualization of vodcasts, debates, anecdotes or stories related to the topic could be told, written or drawn.

From this study, further lines of research could be pursued. Bearing in mind that vodcasts are tools used in mobile language learning, special emphasis might be given to the improvements of audio-visual comprehension not only in formal learning settings, but also in informal ones. Thanks to the proliferation of open-source learning platforms (e. g., Moodle), vodcasts related to the syllabus could be uploaded, as well as audio-visual comprehension activities. These vodcasts could be visualised by language students outside the classroom, deciding when and where their process of learning takes place. The results obtained from the audio-visual comprehension activities completed after watching vodcasts, and recorded in the learning platform, could give us information about the influence of mobile learning on students' foreign language audio-visual comprehension. Another interesting line to follow could be more focused on students' production of vodcasts. Groups of students could create vodcasts with different orchestrations of modes (decided by the teacher or by the students), as well as their audio-visual comprehension activities, which later on other students could watch and complete. This learning situation could be significant since we might determine whether the participation of students in the design of vodcasts is influential in their language learning. Furthermore, we could also explore students' attitudes, motivation and pedagogical implications towards the creation of their own vodcasts and the visualization of vodcasts developed by their partners.

Finally, determining strategies to foment students' audio-visual comprehension is another very important line of research that could benefit, both students and teachers involved in technological educational settings. For example, questionnaires associated to the students' use of metacognitive and cognitive strategies, such as the MALQ (Metacognitive Awareness Listening Questionnaire), created by Vandergrift et al. (2006), could

help us become more familiar with the strategies students use while watching vodcasts. In addition, we could appreciate if there is a need to teach other kinds of strategies for digitally enhanced language learning.

Appendix

A

Test "English is Great"

Name: _____ Age: _____

Comprehension activity 1 - Fill in the missing information with words or numbers from the video.

1. English is the most commonly spoken language across the _____.
2. The library's collection was developed over _____ years.
3. Every year, they have to add _____ kilometres of shelves to house the new item
4. The library has _____million items from all over the world.
5. There are documents that show how _____ changed and evolved over time.
6. A thousand years ago, English was similar to _____.
7. _____ enables people to communicate in English in chat-rooms.
8. Early printers did not make use of _____, dictionaries or guides.

Comprehension activity 2. Answer True or False to the following statements and correct the false ones.

	True/False	Correction
9. The English language is the official language of 45 different countries.		
10. Every year 12 million new items are added in the library.		
11. Shakespeare belongs to the middle English period.		
12. The introduction of lots of technical words produced an impact on the English language.		
13. Early printers had to make up how to spell words themselves.		
14. Roger talks about an essay where the author shortened words.		
15. The English language is very versatile, people play with it and some changes stay.		

TEST "CAMDEN FASHION"

Name: _____ Age: _____

Comprehension activity 1 - Fill in the missing information with words or numbers from the video.

1. In Camden Market there are _____ of stalls, there's something for everyone.
2. Camden is best known for fashion and famous for _____.
3. In the _____ punk fashion shocked many people because it had never been seen before.
4. Punk fashion is still about rejecting _____ fashion.
5. London Fashion week is a great place to see the latest new fashion _____.
6. The shape of the clothes in the British fashion is described as a bit more _____, a bit sharper.
7. A Cyber-Goth is someone who has a very _____ fashion sense.

8. Jack's belt has an influence of the _____.

Comprehension activity 2. Answer True or False to the following statements and correct the false ones.

	True/False	Correction
9. Punks were young people who were anti-government and anti-popular music but loved fashion.		
10. Punk music was loud, wild and about breaking all the rules.		
11. British fashion is quite similar from the fashion that comes from other countries.		
12. British looks are less wearable, less classic and quite boring.		
13. People mix different ideas and create new looks, for instance, Cyber-Goth.		
14. According to Jack, his hair is a logical creation, taking the punk style.		
15. A You-tube top means that it shows the inside of a body.		

B

Test items with answers and the multimodal ensembles orchestrated in the vodcast "*English is GREAT*"

ITEM 1: English is the most commonly spoken language across the globe.
Number of modes: 3
Multimodal ensemble: spoken language, kinesics and music
ITEM 2: The library's collection was developed over two hundred and fifty years.
Number of modes: 3
Multimodal ensemble: spoken language, images and music
ITEM 3: Every year, they have to add twelve kilometres of shelves to house the new items.
Number of modes: 3
Multimodal ensemble: spoken language, images and music
ITEM 4: The library has a hundred and fifty million items from all over the world.
Number of modes: 2
Multimodal ensemble: spoken language and kinesics
ITEM 5: There are documents that show how language changed and evolved over time.
Number of modes: 2
Multimodal ensemble: spoken language and kinesics
ITEM 6: A thousand years ago, English was similar to German.
Number of modes: 2
Multimodal ensemble: spoken language and kinesics
ITEM 7: Technology enables people to communicate in English in chatrooms
Number of modes: 2
Multimodal ensemble: spoken language and kinesics
ITEM 8: Early printers did not make use of grammars, dictionaries or guides.
Number of modes: 2
Multimodal ensemble: spoken language and kinesics
ITEM 9: The English language is the official language of 45 different countries. FALSE (54 countries)
Number of modes: 3
Multimodal ensemble: spoken language, kinesics, music
ITEM 10: Every year 12 million new items are added in the library. FALSE (3 million)
Number of modes: 3
Multimodal ensemble: spoken language, images and music
ITEM 11: Shakespeare belongs to the middle English period. FALSE (early modern English)
Number of modes: 2
Multimodal ensemble: spoken language and kinesics
ITEM 12: The introduction of lots of technical words produced an impact on English language. TRUE
Number of modes: 2
Multimodal ensemble: spoken language and kinesics
ITEM 13: Early printers had to make up how to spell words themselves. TRUE
Number of modes: 2
Multimodal ensemble: spoken language and kinesics
ITEM 14: Roger talks about an essay where the author shortened words. FALSE (a poem)
Number of modes: 2
Multimodal ensemble: spoken language and kinesics

ITEM 15: The English language is very versatile, people play with it and some changes stay. TRUE
Number of modes: 2
Multimodal ensemble: spoken language and kinesics

Test items with answers and the multimodal ensembles orchestrated in the vodcast "*Camden Fashion*"

ITEM 1: In Camden Market there are hundreds of stalls, there's something for everyone.
Number of modes: 3
Multimodal ensemble: spoken language, kinesics and music
ITEM 2: Camden is best known for fashion and famous for punk.
Number of modes: 4
Multimodal ensemble: spoken language, written language, kinesics and music
ITEM 3: In the 1970s punk fashion shocked many people because it had never been seen before.
Number of modes: 3
Multimodal ensemble: spoken language, kinesics and music
ITEM 4: Punk fashion is still about rejecting ordinary fashion.
Number of modes: 4
Multimodal ensemble: spoken language, written language, kinesics and music
ITEM 5: London Fashion week is a great place to see the latest new fashion trends.
Number of modes: 4
Multimodal ensemble: spoken language, written language, images and music
ITEM 6: The shape of the clothes in the British fashion is described as a bit more aggressive, a bit sharper.
Number of modes: 4
Multimodal ensemble: spoken language, written language, kinesics and music
ITEM 7: A Cyber-Goth is someone who has a very strong fashion sense.
Number of modes: 3
Multimodal ensemble: spoken language, written language and kinesics
ITEM 8: Jack's belt has an influence of the 21st century
Number of modes: 4
Multimodal ensemble: spoken language, written language, kinesics and music
ITEM 9: Punks were young people who were anti-government and anti-popular music but loved fashion. FALSE (Anti-fashion)
Number of modes: 4
Multimodal ensemble: spoken language, written language, images and music
ITEM 10: Punk music was loud, wild and about breaking all the rules. TRUE
Number of modes: 4
Multimodal ensemble: spoken language, written language, images and music
ITEM 11: British fashion is quite similar from the fashion that comes from other countries. FALSE (It is different)
Number of modes: 4
Multimodal ensemble: spoken language, written language, images and music
ITEM 12: British looks are less wearable, less classic and quite boring. FALSE (but always very exciting)
Number of modes: 4
Multimodal ensemble: spoken language, written language, kinesics and music
ITEM 13: People mix different ideas and create new looks, for instance, Cyber-Goth. TRUE
Number of modes: 4
Multimodal ensemble: spoken language, written language, images and music
ITEM 14: According to Jack, his hair is a logical creation, taking the punk style. FALSE (a lovely creation)
Number of modes: 3
Multimodal ensemble: spoken language, written language and kinesics
ITEM 15: A You-Tube top means that it shows the inside of a body. FALSE (exo-tube)
Number of modes: 3
Multimodal ensemble: spoken language, written language and kinesics

Notes

This study is based on part of the doctoral thesis presented by the author (Norte, 2016).

Notes

1 At the time of this research in 2014–2015 work on multimodal transcriptions, especially on body movements, was scarce. Since then, other studies such as Querol-Julián and Fortanet-Gómez (2014) have also made use of transcription conventions similar to the ones I have used.

References

Baddeley, A. (2000). The episodic buffer: A new component of working memory? Trends in Cognitive Sciences, 4:417–423.

Baldry, A. P., and Thibault, P. J. (2006). Multimodal Transcription and Text Analysis. London: Equinox.

Chandler, P., and Sweller, J. (1991). Cognitive load theory and the format of instruction. Cognition and Instruction, 8:293–332.

Clark, J. M., and Paivio, A. (1991). Dual coding theory and education. Educational Psychology Review, 3(3):149–170.

Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. System, 29:1–14. doi: 10.1016/S0346-251X(00)00057-9

Firth, J. R. (1957). Papers in Linguistics 1934–1951. London: Oxford University Press.

Goh, C. (2002). Exploring listening comprehension tactics and their interaction patterns. System, 30(2):185–206. doi: 10.1016/S0346-251X(02)00004-0

Goh, C., and Taib, Y. (2006). Metacognitive instruction in listening for young learners. ELT Journal, 60(3):222–232. doi: 10.1093/elt/ccl002

Graham, C. R. (2006). Blended learning systems: Definition, current trends, and future directions. In: Handbook of Blended Learning: Global Perspectives, Local Designs, C. J. Bonk and C. R. Graham (Eds.), 3–21. San Francisco, CA: Pfeiffer Publishing.

Gumperz, J. J., and Berenz, N. B. (1993). Transcribing conversational exchanges. In: Talking Data: Transcription and Coding in Discourse Research, J. A. Edwards and M. D. Lampert (Eds.), 91–121. Hillsdale, NJ: Lawrence Erlbaum.

Halliday, M. A. K., and Hasan, R. (1985). Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective. Oxford: Oxford University Press.

Halliday, M. A. K., and Hasan, R. (1989). Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective. Oxford: Oxford University Press.

Hasan, M., and Hoon, T. B. (2013). Podcasts applications in language learning: A review of recent studies. English Language Teaching, 6(2):128–135. doi: 10.5539/elt.v6n2p12

Hymes, D. H. (1974). Foundations in Sociolinguistics: An Ethnographic Approach. Philadelphia: University of Pennsylvania Press.

Jefferson, G. (1984). On stepwise transition from talk about trouble to inappropriately next-positioned matters. In: Structures of Social Action: Studies in Conversation Analysis, J. M. Atkinson and J. Heritage (Eds.), 191–221. Cambridge: Cambridge University Press.

Jewitt, C. (2013). Multimodal methods for researching digital technologies. In: The SAGE Handbook of Digital Technology Research, S. Price, C. Jewitt and B. Brown (Eds.), 250–265. London: Sage.

Jewitt, C., and Kress, G. (Eds.) (2003). Multimodal Literacy. New York: Peter Lang.

Kress, G. (1993). Against arbitrariness: The social production of the sign as a foundational issue in critical discourse analysis. Discourse and Society, 4(2):169–191. doi: 10.1177/0957926593004002003

Kress, G. (2010). Multimodality: A Social Semiotic Approach to Contemporary Communication. London: Routledge.

Kress, G., and van Leeuwen, T. (2001). Multimodal Discourse: The Modes and Media of Contemporary Communication. London: Edward Arnold.

Malinowski, B. (1923). The problem of meaning in primitive languages. In: The Meaning of Meaning. Supplement I, C. K. Ogden and I. A. Richards (Eds.), 296–336. New York: Harcourt Brace & World.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In: The Cambridge Handbook of Multimedia Learning, R. E. Mayer (Ed.), 31–48. New York: Cambridge University Press.

Mayer, R. E. (2009). Multimedia Learning. New York: Cambridge University Press.

Mendelsohn, D. J. (1994). Learning to Listen: A Strategy-Based Approach for the Second Language Learner. San Diego: Dominie Press.

Mendelsohn, D. J., and Rubin, J. (Eds.). (1995). A Guide for the Teaching of Second Language Listening. San Diego, CA: Dominie Press.

Morell, Teresa. (2018). Multimodal competence and effective interactive lecturing. System, 77: 70-79./10.1016/j.system.2017.12.006.

Norris, S. (2004). Analyzing Multimodal Interaction: A Methodological Framework. New York & London: Routledge.

Norris, S. (2009). Modal density and modal configuration: Multimodal actions. In: The Routledge Handbook of Multimodal Analysis, C. Jewitt (Ed.), 78–90. London: Routledge.

Norris, S., and Jones, R. H. (Eds.). (2005). Discourse in Action: Introducing Mediated Discourse Analysis. London: Routledge.

Norte Fernández-Pacheco, N. " *The orchestration of modes and EFL audio-visual comprehension: A multimodal discourse analysis of vodcasts.*". Spain: University of Alicante, 2016. Department of English Studies, unpublished PhD thesis.

O'Malley, J. M., and Chamot, A. U. (1990). Learning Strategies in Second Language Acquisition. Cambridge: Cambridge University Press.

Oxford, R. L. (1990). Language Learning Strategies: What Every Teacher Should Know. Boston, MA: Heinle & Heinle.

Plonsky, L., and Osward, F. (2014). How Big Is «Big»? interpreting effect sizes in L2 research. Language Learning, 64(4):878–912. doi: 10.1111/lang.12079

Prensky, M. (2001). Digital Natives, Digital Immigrants Part I. On the Horizon, 9(5):1–6. doi: 10.1108/10748120110424816

Querol-Julián, M., and Fortanet-Gómez, I. (2014). Theoretical framework for a multimodal analysis of evaluation in discussion sessions of conference paper presentations. Kalbotyra, 66:77–98.

Ramirez, D., and Alonso, I. (2007). Using digital stories to improve listening comprehension with Spanish young learners of English. Language Learning & Technology, 11(1):87–101. http://goo.gl/U5z8S0 (2016-05-19).

Rosell- Aguilar, F. (2007). Top of the pods- in search of a podcasting «podagogy» for language learning. Computer Assisted Language, 20(5):471–492. doi: 10.1080/09588220701746047
Rubin, J. (1994). A review of second language listening comprehension research. The Modern Language Journal, 78:199–221. doi: 10.1111/j.1540-4781.1994.tb02034.x
Sloetjes, H., and Wittenburg, P. (2008). *Annotation by category - ELAN and ISO DCR*. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008).
Sueyoshi, A., and Hardison, D. (2005). The role of gestures and facial cues in second language listening comprehension. Language Learning, 55:661–699. doi: 10.1111/j.0023-8333.2005.00320.x
Suvorov, R. (2009). Context visuals in L2 listening tests: The effects of photographs and video vs. audio-only format. In: Developing and Evaluating Language Learning Materials, C. A. Chapelle, H. G. Jun and I. Katz (Eds.), 53–68. Ames: Iowa State University.
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction, 4(4):295–312.
Tan, S. (2009). A systemic functional framework for the analysis of corporate television advertisements. In: The World Told and the World Shown: Multisemiotic Issues, E. Ventola and A. J. M. Guijarro (Eds.), 157–182. Hampshire: Palgrave Macmillan.
Vandergrift, L. (1999). Facilitating second language listening comprehension: Acquiring successful strategies. ELT Journal, 53(3):168–176. doi: 10.1093/elt/53.3.168
Vandergrift, L. (2003). Orchestrating strategy use: Towards a model of the skilled L2 listener. Language Learning, 53:461–494. doi: 10.3138/cmlr.59.3.425
Vandergrift, L. (2008). Learning strategies for listening comprehension. In: Language Learning Strategies in Independent Settings, S. Hurd and T. Lewis (Eds.), 84–102. Clevedon, GBR: Channel View Publications.
Vandergrift, L., and Goh, C. (2012). Teaching and Learning Second Language Listening: Metacognition in Action. New York & London: Routledge.
Vandergrift, L., Goh, C., Mareschal, C., and Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire (MALQ): Development and validation. Language Learning, 56(3):431–462. doi: 10.1111/j.1467-9922.2006.00373.x
Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. Language Testing, 27:493–513. doi: 10.1177/0265532209355668

Bionotes

Natalia Norte Fernández-Pacheco is an adjunct lecturer at the University of Alicante (UA) and an English teacher at CESA (Alicante Superior Studies Centre). She obtained her degree in English Studies thanks to her studies at the University of Alicante, as well as at the National University of Ireland, Maynooth. Currently, her professional work at CESA is focused on teaching English to both children and adults, immersed in different levels of the Common European Framework of Reference for Languages. At the same time, she teaches in the Department of English Studies (UA) in subjects of the Degrees in Tourism and in English Studies.With reference to her research career, this is based on the analysis of multimodal discourse and the teaching of English as a foreign language. Likewise, this line of research is complemented by the study of new technologies applied to English teaching. In February 2007, she obtained the diploma of advanced studies (DEA), with the research work entitled "The effects of multimodal presentations in EFL content university lectures", directed by Dr. Teresa Morell Moll. In January 2016, she defended the thesis "The orchestration of modes and EFL audio-visual comprehension: A multimodal discourse analysis of vodcasts", also supervised by Dr. Teresa Morell Moll. Dr. Natalia Norte is part of the ACQUA research group (Research in second and foreign language acquisition at the University of Alicante), directed by Dr. Susana Pastor Cesteros. https://cvnet.cpd.ua.es/curriculum-breve/es/norte-fernandez-pacheco-natalia/81121